

About Space and Time Dimensions of Enterprise Data

Prof.dr. Constantin-Gelu APOSTOL
Catedra de Informatică Economică, A.S.E. București

Time and space are discussed as main dimensions of the information and, consequently, of the different kinds of enterprise data. Some theoretical approaches on spatiotemporal reasoning are considered and, based on it, a few practical consequences are identified, especially for the enterprise data warehouse and data marts. The final part investigates the main characteristics of temporal object models.

Keywords: enterprise data, enterprise application, space dimension, time dimension, spatiotemporal reasoning, temporal object model.

Introducere

Aplicațiile informatice de întreprindere se caracterizează prin prelucrarea unor volume mari de date, specifice diverselor activități de afaceri. Datele sunt identificate prin nume și corespund unor valori (sau mulțimi de valori) reprezentate în formă digitală, cu **dimensiuni** în **timp** și **spațiu** definite prin attribute ale tipurilor de date folosite.

Reprezentarea și manipularea riguroasă a relației dintre laturile spațială și temporală ale datelor sunt condiții esențiale pentru funcționarea și integrarea aplicațiilor informatice, având în vedere schimbările permanente din activitatea internă și din mediul cu care întreprinderile interacționează. De altfel, frecvent se folosește conceptul de **continuum spațio-temporal**, ca ilustrare a interdependențelor dintre cele două laturi ale oricărei activități, fenomen sau proces din lumea reală, cu preocupări de extindere și la activitățile virtuale ale spațiului cibernetic (Grosbeck, 2003).

Spațiul și timpul ca dimensiuni ale informației

Timpul și spațiul sunt attribute esențiale ale informației, obligatorii pentru reflectarea corectă a lumii înconjurătoare în permanentă schimbare. Corespunzător, datele, ca model de reprezentare a informației, trebuie să asigure o reprezentare adecvată nu numai a valorii acesteia, dar și a dimensiunilor spațio-temporale asociate.

Folosind notația formală BNF (*Backus-Naur Form*), data poate fi definită în forma (Apostol, 2003):

$$\langle \text{dată} \rangle ::= \langle \text{nume} \rangle \langle \text{attribute} \rangle \langle \text{valoare} \rangle \quad (1)$$

Dimensiunile de timp și spațiu, ca și alte caracteristici (natură, reprezentare internă etc.) sunt definite prin $\langle \text{attribute} \rangle$ care, împreună cu numele datei, se constituie într-un **lanț de identificatori** definind univoc fiecare dată.

Chiar în condițiile în care, aparent, caracteristicile de spațiu rămân neschimbate, curgea timpului obligă la înregistrări, uneori continue, dar adesea discontinue (discrete) ale valorilor datelor, ajungându-se astfel la serii de date. Frecvența de înregistrare depinde de variabilitatea în timp a caracteristicilor entităților reflectate informațional prin date, în care simpla distincție între **date variabile** și **date constante** nu mai este suficientă.

Referitor la prima categorie, este evidentă existența unor grade diferite de variabilitate, determinate obiectiv de dinamica evoluției caracteristicilor reprezentate. Totodată, trebuie subliniată și distincția între dinamica reală și periodicitatea observării și înregistrării acestor date, care din rațiuni economice sau limitări tehnice poate fi diferită.

Cât privește a doua categorie, o denumire mai potrivită ar fi aceea de **date convențional constante**, având în vedere că și ele se pot schimba în timp, dar în condiții imprevizibile.

Abordări teoretice și implicații practice ale dimensiunilor spațio-temporale

Cercetările privind **sistemele informaționale geografice** (*GIS – Geographic Information Systems*), destinate reprezentării și prelucrării datelor geo-spațio-temporale, au avut un rol de prim rang în aprofundarea dimensiunilor de spațiu și timp ale informației (Ozsu et al.,

1996), (Liou, 1999), (Bittner & Smith, 2003), (Grenon & Smith, 2004) etc.

Pentru a da un singur exemplu, menționăm lucrarea publicată de Grenon și Smith (2004), în domeniul ontologiei formale, din care considerăm relevant următorul citat:

“Realitatea este esențial dinamică. În fața acestui fapt, se poate distinge între două abordări alternative, tradițional percepute ca incompatibile. Pe de o parte, există abordări care iau foarte în serios pe plan ontologic dihotomia între spațiu și timp. Pe de altă parte, există abordări care percep realitatea în termenii unei singure teorii ontologice unificate spațio-temporale. Aceste două viziuni corespund la două forme diferite de raționamente spațio-temporale. Prima lucrează cu succesiuni de reprezentări instantanee (snapshots) ale lumii înconjurătoare, ultima cu schimbările și procesele ca atare”.

Demersul autorilor citați are în vedere definierea unei ontologii care să înglobeze ceea ce este valoros în ambele abordări, abordând realitatea spațială atât **sincronic** (așa cum există la un moment dat), cât și **diacronic** (așa cum se derulează în timp). Corespunzător, sunt definite formal două teorii ontologice, SNAP (de la *snapshot*), respectiv SPAN (denumire aleasă prin contrast cu prima), corespunzând celor două abordări evidențiate și se demonstrează posibilitatea reunirii lor într-o viziune unitară.

Pe planul unor instrumente și tehnologii informatice care să permită aplicarea practică a unor asemenea ontologii, poate fi remarcată dezvoltarea unor noi generații de sisteme de gestiune a bazelor de date (SGBD), între care se detașează cele orientate obiect (SGBD-OO), oferind mecanismele pentru modelarea și manipularea unor obiecte de o complexitate dimensională spațio-temporală arbitrară. În timp ce modelul relațional obligă la descompunerea reprezentării unui obiect complex într-o mulțime de relații, abordările orientate obiect permit utilizatorilor să imite și să simuleze lumea reală într-o manieră flexibilă, beneficiind de avantajul reutilizării. Aplicațiile în care dimensiunea temporală este esen-

țială pot beneficia de suportul oferit de SGBD-OO și în surprinderea mai bună a semnificativității obiectelor complexe care survin în diferite domenii de activitate, cum ar fi cel al GIS.

Din perspectiva modelării și prelucrării dimensiunii timp, se cuvine remarcată constituirea ca direcție distinctă de cercetare a sistemelor de **baze de date temporale** (*temporal databases*), cu fundamentarea teoretică a extensiilor în raport cu SGBD convenționale, dar și cu metodologii de proiectare la nivel conceptual, logic și fizic, respectiv instrumente CASE pentru asistarea diverselor etape ale acestui proces (Detienne & Hainaut, 2001).

Reprezentarea unitară și gestiunea riguroasă a dimensiunii temporale a datelor depășește granițele diverselor activități, atunci când se trece la crearea de **depozite de date** (*data warehouse*), resursă esențială în sistemele de asistare a deciziilor, în implementarea aplicațiilor de inteligență a afacerilor (BI - *Business Intelligence*) sau pentru integrarea aplicațiilor din întreprindere (EAI - *Enterprise Application Integration*).

Abordarea uzuală în construirea depozitului de date central este una incrementală, bazată pe construirea progresivă a unor **depozite departamentale** (*data mart*) care se alimentează din diverse surse de date – adesea eterogene - și pot conține date mai sintetice și mai puțin detaliate decât depozitul central (figura 1). Așa cum evidențiază figura 1, realizarea acestor transferuri succesive este condiționată de asigurarea consistenței spațio-temporale între diversele clase de entități implicate, cu eventuale conversii pentru satisfacerea acestei cerințe.

Această abordare poate fi convenabilă pentru utilizatorii finali care își concentrează analizele datelor pe domenii specifice din activitatea întreprinderii.

Pe de altă parte, așa cum evidențiază Halpin (2001), pentru interogările analitice care se referă la domenii multiple, devine critică o viziune unitară de ansamblu, construită pe baza modelului de date al întreprinderii.

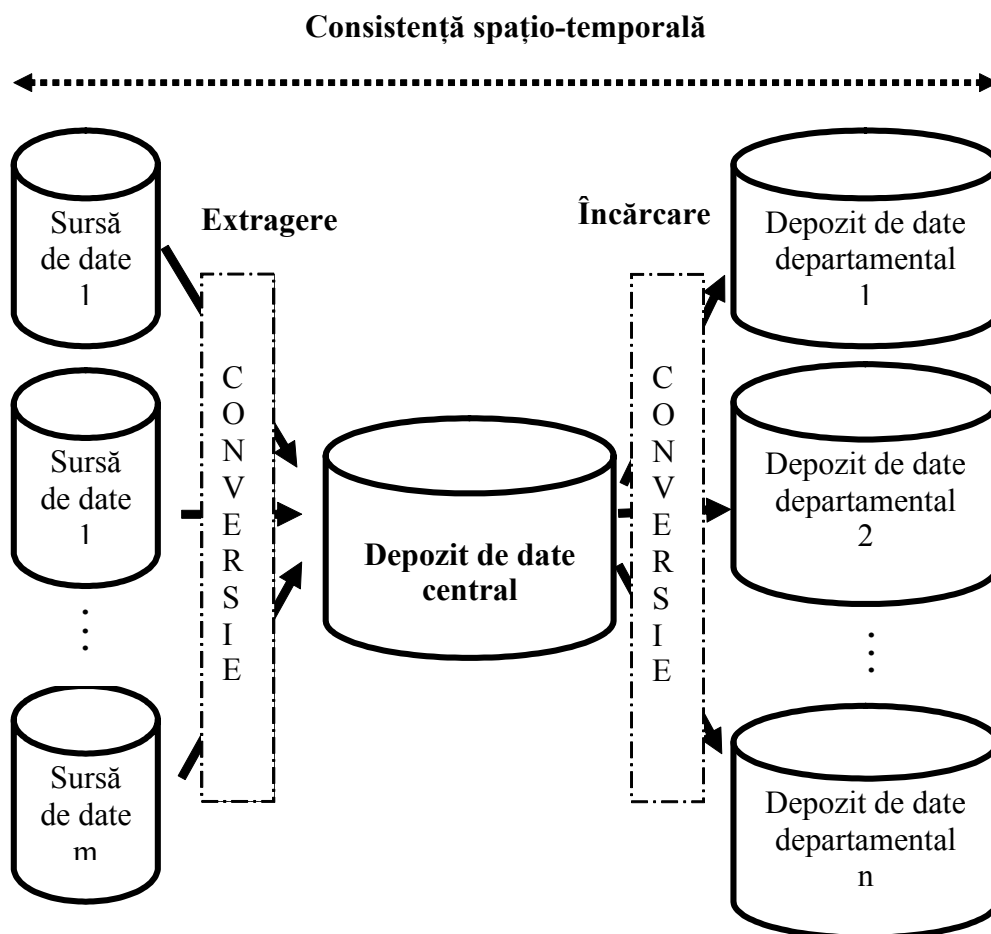


Fig. 1. Extragerea datelor incluse în depozitul central și încărcarea în depozitele departamentale

Modele obiect temporale

Proiectanții de aplicații și de depozite de date au în vedere că datele factuale (referitoare, de exemplu, la comenzi de la clienți sau livrări de produse) se acumulează în ritm rapid. În același timp, atributele folosite ca suport al dimensiunilor acestor date, cum sunt numele clientului sau mărimea produsului sunt comparativ statice (sau, cu un termen propus anterior – convențional constante). Chiar dacă multe dintre dimensiuni pot fi subiectul unor schimbări lente (referite uneori cu acronimul SCD – *Slowly Changing Dimension*), proiectanții au obligația de a include funcționalități destinate acestui scop.

Abordările elementare sunt denumite generic ca fiind de tipul 1, 2 sau 3 (Ross & Kimball, 2005):

1. **Tipul 1** pare cel mai potrivit în cazul realizării corecțiilor, dar această tehnică nu va conserva cu acuratețe schimbările în timp ale valorilor atributelor. Atributul schimbat va fi

pur și simplu suprascris, reflectând întotdeauna numai valoarea curentă.

2. În cazul schimbărilor de **tipul 2**, păstrarea istoricului este asigurată prin includerea în tabela atributului (dimensiunii) respectiv a unei noi linii (tuplu), cu o cheie primară surrogat. Atît linia veche cît și cea nouă conțin ca atribute cheia naturală (sau identificatorul durabil), precum și datele calendaristice de intrare în vigoare și expirare.

3. Pentru **tipul 3**, liniei existente a dimensiunii îi este asociat un alt atribut, pentru a face posibilă analizele bazate fie pe valoarea nouă, fie pe cea anterioară a atributului.

Fără a ne putea propune în acest cadru o analiză de detaliu a utilizării acestor soluții, remarcăm doar că situațiile survenite în practică pot conduce la necesitatea unor **variante hibride**, prin diverse combinații sau extensii ale celor trei tipuri de bază. Implementarea efectivă acestor mecanisme va depinde, totodată, de funcționalitățile oferite de SBGD-

urile utilizate.

Abordări de tipul celei prezentate evidențiază necesitatea de a surprinde cât mai complet fațetele atributului timp, ceea ce a condus la apariția mai multor propuneri de modele obiect temporale, între care se constată unele similitudini semnificative (Özsu et al., 1996):

- Fiecare model are una sau mai multe primitive temporale, adică moment de timp (*time instant*), interval de timp (*time interval*), perioadă de timp (*time span*), timp continuu și discret etc.;
- Anumite modele impun ca primitivele lor temporale să aibă aceeași granularitate de bază, în timp ce altele acceptă granularități multiple și permit ca primitivele temporale să fie exprimate în diferite granularități;
- Cele mai multe modele temporale acceptă un model liniar și ciclic al timpului, în timp ce unele acceptă un model ramificat;
- Toate modelele temporale oferă anumite mijloace de modelare a informației istorice a entităților lumii reale și a istoricului entităților din baza de date. Două dintre cele mai cunoscute metode de modelare a istoricului entităților sunt timpul de validitate (*valid time history*) și timpul de tranzacționare (*transaction time history*).

Fără a ne propune, în acest cadru, o analiză a diverselor modele, remarcăm importanța pentru aplicabilitatea lor a semnificației asociate conceptului de **granularitate**, care stă apoi la baza definirii celui de **primitivă temporală**.

Pe linia creării unei abordări unitare, se cuvine semnalată soluția propusă în (Goralwalla et al., 2001). Mai întâi este necesară distincția între **datele temporale ancorate** (de exemplu, *1 ianuarie 2006* sau *ianuarie 2006*) și cele **neancorate** (de exemplu, *30 secunde*, *5 ore 15 minute*, *7 zile*, *3 luni* etc.).

În practică asigurarea comparabilității datelor și a utilizării acestora pentru diverse analize presupune conversia lor de la o granularitate la alta (vezi, de exemplu, figura 1). Evidențind că majoritatea soluțiilor propuse au în vedere conversiile de granularitate pentru datele temporale ancorate, autorii propun în lucrarea referită o nouă abordare a granularității pentru datele temporale. Astfel, granularitatea este modelată ca un tip special

de primitivă temporală neancorată, care poate fi apoi folosită ca o unitate de timp. Se creează astfel o bază unitară pentru reprezentarea datelor temporale neancorate sau ancorate, pentru crearea procedurilor de conversie a datelor de la o granularitate la alta, precum și pentru stabilirea modului de realizare a operațiilor între date.

Bibliografie

1. Apostol, C.G. (2003). *Informatică pentru afaceri (partea I-a)*, Editura ASE, București
2. Bittner, T., Smith, B. (2003). *Granular Spatio-Temporal Ontologies*, în: 2003 AAAI Symposium: Foundations and Applications of Spatio-Temporal Reasoning (FASTR), AAAI Press, pag. 12-17
3. Detienne, V., Hainaut, J.L. (2001). *Case Tool Support for Temporal Database Design*, <http://www.info.fundp.ac.be/~dbm/publication/2001/er2001-55.pdf>
4. Grenon, P., Smith, B. (2004). *SNAP and SPAN: Towards Dynamic Spatial Ontology*, în: *Spatial Cognition and Computation*, 4 (2004) 1, pag. 69-104
5. Goralwalla, I.A., Leontiev I., Özsu, M.T., Szafron, D., Combi, C. (2001). *Temporal Granularity: Completing the Puzzle*, *Journal of Intelligent Information Systems*, Vol.16, Nr.1, pag. 41-63
6. Grossecck, G. (2003). *Incursiune în ciberspațiu*, 2003, în *Informatica economică*, nr.4(28), pag. 9-14
7. Halpin, T. (2001). *Information Modeling and Relational Databases – From Conceptual Analysis to Logical Design*, Morgan Kaufmann Publishers, San Francisco
8. Liou, J. (1999). *Temporal Support for Land Information Systems in Object-Oriented Modeling*, Doctoral Thesis, Royal Institute of Technology, Stockholm, <http://www.lib.kth.se/Fulltext/liou991220.pdf>
9. Özsu, M.T., Szafron, D., Goralwalla, I.A. (1996). *Modeling time: Back to basics*. TR-96-03, Dept. of Computer Science, University of Alberta.
10. Ross, M., Kimball R. (2005). *Slowly Changing Dimensions Are Not Always as Easy as 1, 2, 3*, *Intelligent Enterprise*, http://www.intelligententerprise.com/news_an_jhtml